Learner-adaptive partial and synchronized captions for L2 listening skill development

Maryam Sadat Mirzaei[1] and Kourosh Meshgi[2]

Abstract. Many language learners have difficulty practicing listening skills using authentic materials, and thus use captions to map text with speech, and they benefit from reading along while listening to comprehend content. However, many learners over-rely on reading the text and many have difficulty in dividing their attention to the multimodal input. We have proposed a captioning tool, Partial and Synchronized Captions (PSC), which detects the useful words to be shown in the caption for addressing learners' listening difficulties. To handle individual learner demands, PSC should adapt its word selection criteria. This study proposes an Adaptive PSC (APSC), which improves its word selection and retrains itself on-the-fly by applying learner feedback on the generated caption to provide individualized and effective assistance that satisfies the learners' requirements. Preliminary results revealed that the system was relatively successful to adapt itself to the demand of the L2 learner, which raised learner satisfaction on the resultant captions.

Keywords: partial and synchronized captions, adaptive captions, individualized assistance, second language listening.

1.   Introduction

One popular tool used for developing L2 listening skills, especially when it comes to listening to authentic materials, is captioning (Vandergrift, 2011). Captioning facilitates listening comprehension by providing the text along with the audio/video. However, many learners, especially beginners, struggle with cognitive load and split attention, while attending to caption text together with other modes of input (Leveridge & Yang, 2013; Sweller, 1994). Mirzaei, Meshgi, Akita, and

Kawahara (2017) proposed PSC to provide selective text in the caption for reducing textual density and encouraging more listening than reading. PSC synchronizes text and audio at word-level to facilitate text-to-speech mapping. The selection of words to appear in the caption is based on lexical and speech difficulty. The former considers factors such as frequency and specificity, whereas the latter incorporates the use of automatic speech recognition on the system's errors to detect difficult speech segments (e.g. breached boundaries).

The main challenge is the word selection for learners with different proficiencies. While the full caption may bring too much text that sometimes negatively affects the comprehension (Leveridge & Yang, 2013), partial captioning may provide insufficient text for beginners or too much text for highly-advanced learners. One solution is to make an interactive environment where learners can provide feedback to the system on selected words. Meanwhile, the system should be able to learn from learners' feedback to address individual's needs.

This paper proposes a machine learning approach that uses learner's feedback on-the-fly to adapt the word selection criteria of PSC with the ever-changing user preferences and video stream. Therefore, we asked the learners to mark the hidden words that they wanted to see in PSC and to omit shown words that were easy for them. The system is then trained based on the learner's feedback and adapted its word selection accordingly (Figure 1).

Figure 1.   Learner feedback on the caption to hide a word (top) and to show a masked word (bottom) by clicking on it. The classifier's labels and the decision boundary (dashed line) change according to the learner feedback
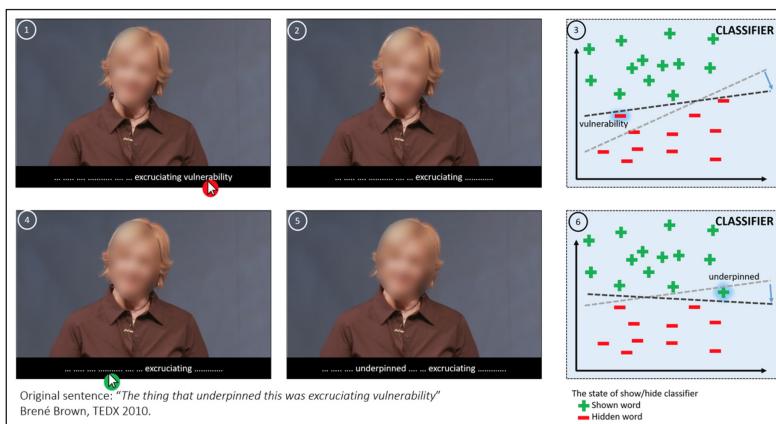
2. APSC

Different lexical, acoustic, and content-based features are considered in PSC. The features are extracted for each word, classifying it as either easy or difficult. A word is classified as difficult when its feature value exceeds some thresholds. Mirzaei et al. (2017) proposed using learners' vocabulary and listening test scores to adjust the thresholds for filtering words and making a caption for learners of similar proficiency groups. However, such a method ignores individual differences within each proficiency group, the limitation of the tests to measure the different listening difficulty features, and the effect of learners' background on their listening comprehension (e.g. engineers listening to medical talks). Moreover, the fixed threshold does not reflect the gradual improvement of learner's listening skills.

Previous analyses revealed that some learners need additional factors to be considered when generating PSC (e.g. speech disfluencies) and others gradually adapt to the listening material (e.g. getting used to vocabulary and speech rate of the speaker), hence no longer needing some words in the caption. To this end, we developed the APSC (Figure 2), in which an online machine learning module receives the feedback from the learners and adjusts the thresholds of the system on-the-fly. The feedback includes user clicks either on a masked word they wish to see or on a shown word that is too easy. The system reacts by showing/hiding the word and learns to intelligently classify words with similar features in the future.

Figure 2.   System components and process flow: TED talks are fed to the forced aligner for text-to-speech synchronization. Knowledge bases (e.g. corpora) are used for detecting listening difficulty features. The classifier applies the rules and learner feedback for word selection
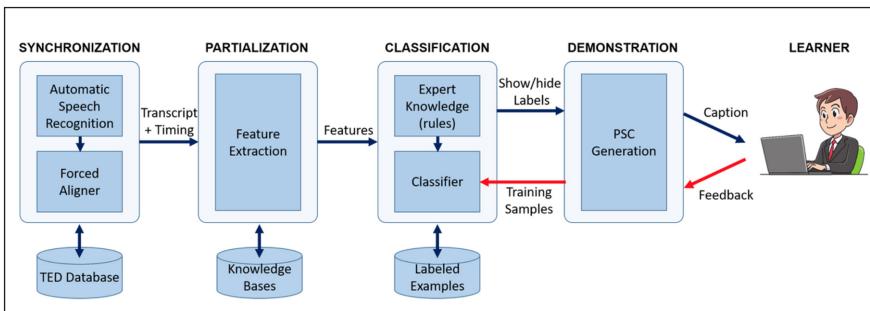
Rather than defining rules, our classifier is trained by giving several examples for each category of words in context. Therefore, it can easily expand to support other

types of listening materials (e.g. daily conversations, news) that require different rules, features, and thresholds. Additionally, the system can detect and learn the discriminative features of the learners' feedback. Such feedback serves as a bag of examples for retraining the system, which can be easily obtained from the learners and used to their advantage. The learner feedback acts as new labels for words that the system misclassified, and the classifier is re-trained with such data to learn about individuals' problems, backgrounds, vocabulary reservoirs, and possible sources of listening difficulties.

3.   Preliminary evaluation and discussion

Twenty-four pre-intermediate learners of English, graduate students of Kyoto University with engineering and humanities backgrounds, used our system and provided feedback. They were divided into two groups and were asked to: (1) watch a series of videos captioned by using baseline PSC and provide feedback by clicking on difficult words masked (to be shown), and on the easy words that were shown (to be hidden); and (2) watch another set of videos and provide feedback in a similar way, however, this time the first group received baseline PSC (i.e. their feedback was received but not applied in the PSC), whereas the second group received APSC trained by their annotations in the previous phase.

For each set, learners were given five different two to three minute TED Talk clips delivered by native English speakers. Learners were also asked to do a five-point Likert scale questionnaire on the use of system.

Analysis of the number of modifications for the first and second sets of videos revealed that learners who received APSC required fewer modifications in the second round (M=14.2, SD=1.6) compared to those whose feedback was not applied (M=9.8, SD=2.0). The difference was statistically significant, [$t(8)$=3.74; p=0.006], indicating that the group who received APSC were generally more satisfied with the captions generated by the trained system and required fewer modifications (Figure 3).

Learner feedback on the questionnaire (Figure 4) demonstrated that they enjoyed having control over the captions (Q1, Q4), benefited from individualized captions (Q3, Q5), and were motivated to use the system (Q6) with less frustration (Q7, only asked from the second group). Most learners also believed that this system can be more interesting and useful if it challenges them with more difficult cases (Q2).

Detailed analysis revealed that words with different British and American pronunciations were selected more frequently to be included in the caption. The learners also demanded to show idioms and sentences with complex grammar. Moreover, talks delivered by specific speakers raised more feedback, perhaps due to many speech disfluencies. Additionally, learners with certain backgrounds chose to hide certain words in the captions that were familiar to them.

This system aims to overcome the shortcomings of keyword or partial captioning on ignoring different learner's requirements (Guillory, 1998). Furthermore, unlike the full caption, this system reduces text to facilitate ingesting the multimodal input (Vandergrift, 2011), provides learner control over the generated captions, and tailors the captions for different learners to increase satisfaction.

Figure 3. Analyzing the number of modifications requested in the captions by the learners as an indicator of learner satisfaction of word selection in PSC
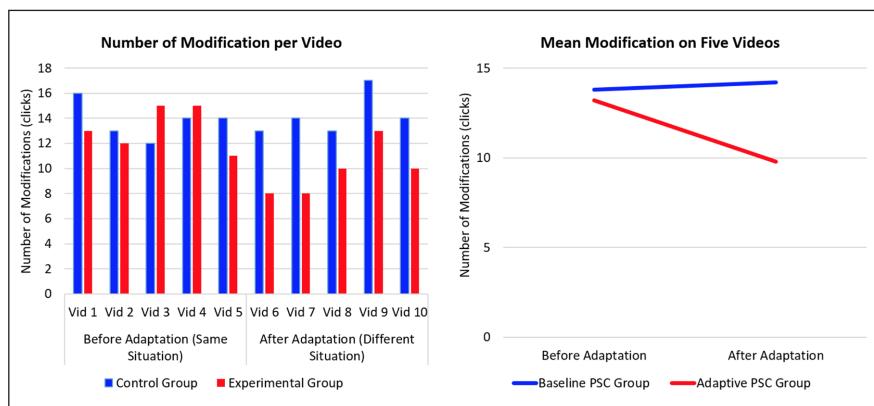
Figure 4. Learner feedback on the questionnaire (APSC group only)
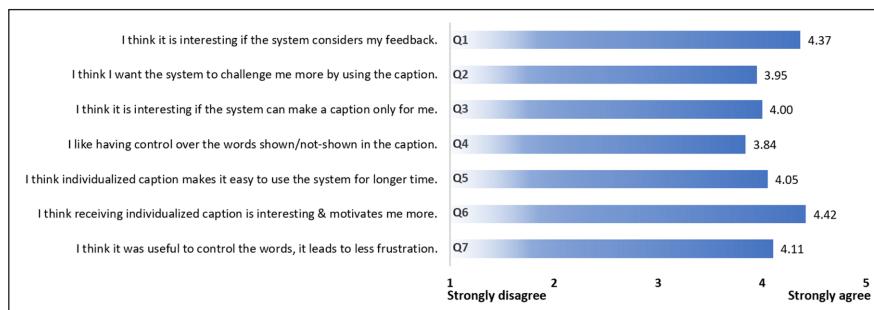
4.   Conclusions

We developed an APSC system that considers learner feedback on the word selection, trains itself based on such feedback, and provides more individualized captioning for each learner. The system uses machine learning to identify the listening difficulties of learners by using their feedback as example cases and provides effective scaffolding by selecting necessary words for the captions. System evaluation revealed that our approach is successful in providing tailored captions to the listeners, thus increasing learner satisfaction, while the effectiveness of the system largely depends on the amount of feedback each learner provides.

5.   Acknowledgments

References

Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *Calico Journal*, *15*(1-3), 89-108. https://doi.org/10.1558/cj.v15i1-3.89-108

Leveridge, A., & Yang, J. (2013). Testing learner reliance on caption supports in second language listening comprehension multimedia environments. *ReCALL*, *25*(2), 199-214. https://doi.org/10.1017/s0958344013000074

Mirzaei, M. S., Meshgi, K., Akita, Y., & Kawahara, T. (2017). Partial and synchronized captioning: a new tool to assist learners in developing second language listening skill. *ReCALL*, *29*(2), 178-199. https://doi.org/10.1017/s0958344017000039

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295-312. https://doi.org/10.1016/0959-4752(94)90003-5

Vandergrift, L. (2011). Second language listening: presage, process, product, and pedagogy. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (455-471). Routledge.

CALL and complexity – short papers from EUROCALL 2019
Edited by Fanny Meunier, Julie Van de Vyver, Linda Bradley, and Sylvie Thouësny